

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:42:28 ; Search time 0.512821 Seconds
(without alignments)
1299.194 Million cell updates/sec

Title: US-09-360-685B-10

Perfect score: 27

Sequence: 1 EPIYA 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_101002:*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
 - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
 - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
 - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
 - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
 - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
 - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
 - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
 - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
 - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
 - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
 - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
 - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
 - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
 - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
 - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
 - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
 - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
 - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
 - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
 - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
 - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
 - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	17	23	AAE19658 Human tumour necro
2	27	100.0	47	22	ABB42462 Peptide #9968 enco
3	27	100.0	47	22	ABB25897 Protein #7896 enco
4	27	100.0	47	22	AA63351 Human brain expres
5	27	100.0	47	22	AA676165 Human bone marrow
6	27	100.0	47	22	AA20819 Peptide #7253 enco
7	27	100.0	47	22	AA36275 Peptide #10312 enc
8	27	100.0	47	23	ABG45519 Human peptide enco
9	27	100.0	85	18	AAW23007 Canine herpesvirus
10	27	100.0	85	19	AAW72669 Canine herpes viru

11	27	100.0	85	22	AAE19658	Canine herpes viru
12	27	100.0	93	22	AAE19658	C glutamicum prote
13	27	100.0	115	21	AAE19658	Human signal pepti
14	27	100.0	115	23	ABB89722	Human polypeptide
15	27	100.0	118	23	AAE19658	Human ghltnF40.4
16	27	100.0	118	23	AAE19658	Human ghltnF40.4
17	27	100.0	150	21	AAE19658	Human membrane-ass
18	27	100.0	181	22	ABG01845	Novel human diagno
19	27	100.0	207	22	ABG13524	Novel human diagno
20	27	100.0	216	23	AAE19658	Antibody Fab fragm
21	27	100.0	229	23	AAE19658	Antibody modified
22	27	100.0	244	23	ABP45502	Human BlyS binding
23	27	100.0	338	22	AAE19658	Follicular conjunc
24	27	100.0	338	22	AAE19658	Follicular conjunc
25	27	100.0	376	19	AAW82623	Ehrlichia sp. HGP-
26	27	100.0	376	21	AAE19658	Ehrlichia antigen
27	27	100.0	376	23	AAU73193	Human granulocytic
28	27	100.0	449	22	ABB60974	Drosophila melanog
29	27	100.0	451	22	ABB60974	Drosophila melanog
30	27	100.0	506	22	ABB62754	Drosophila melanog
31	27	100.0	573	20	AAW82490	Ehrlichia sp. E82
32	27	100.0	712	23	AAU73232	Human granulocytic
33	27	100.0	721	21	AAE19658	Granulocytic ehrl
34	27	100.0	721	21	AAE19658	Granulocytic ehrl
35	27	100.0	722	21	AAE19658	Granulocytic ehrl
36	27	100.0	739	21	AAE19658	Granulocytic ehrl
37	27	100.0	747	21	AAE19658	Granulocytic ehrl
38	27	100.0	747	21	AAE19658	Granulocytic ehrl
39	27	100.0	747	21	AAE19658	Granulocytic ehrl
40	27	100.0	747	21	AAE19658	Granulocytic ehrl
41	27	100.0	747	21	AAE19658	Granulocytic ehrl
42	27	100.0	747	21	AAE19658	Granulocytic ehrl
43	27	100.0	748	20	AAW89272	Granulocytic ehrl
44	27	100.0	748	21	AAE19658	Granulocytic ehrl
45	27	100.0	748	21	AAE19658	Granulocytic ehrl

ALIGNMENTS

RESULT 1
AAE19658

ID AAE19658 standard; peptide; 17 AA.

XX AAE19658;

XX DT 31-MAY-2002 (first entry)

XX DE Human tumour necrosis factor (TNF) 40 hybrid CDRH2' peptide.

XX KW Human; tumour necrosis factor 40; TNF40; osteopathic; cardiant; CDR;
KW complementarity determining region; rheumatoid; osteo-arthritis; sepsis;
KW congestive heart failure; shock; tissue transplant; tuberculosis; AIDS;
KW Acquired Immune Deficiency Syndrome; adult respiratory distress syndrome;
KW cachexia; allergy; psoriasis; blood coagulation disorder; thyroiditis;
KW inflammatory bone disorder; Crohn's disease; autoimmune disease; burn;
KW neoplasia therapy; immunomodulator; vulnery; graft rejection.

XX OS Homo sapiens.

XX PN WO200194585-A1.

XX PD 13-DEC-2001.

XX PF 05-JUN-2001; 2001WO-GB02477.

XX PR 06-JUN-2000; 2000GB-0013810.

XX PA (CELL-) CELLTECH R & D LTD.

XX PI Athwal DS, Brown DT, Weir ANC, Popplewell AG, Chapman AP, King DJ;

XX WPI; 2002-216732/27.

XX New antibody specific for human tumor necrosis factor (TNF)-alpha,
PT useful for treating TNF-alpha-mediated diseases, e.g. congestive heart
PT failure, septic or endotoxic shock, cachexia, adult respiratory
PT distress syndrome
XX
XX Claim 1; Fig 3; 119pp; English.
XX
CC The invention relates to an antibody molecule having specificity for
CC human tumour necrosis factor-alpha (TNFalpha) comprising a heavy or light
CC chain. The antibody or the compound comprising the antibody is useful
CC for treating or manufacturing a medicament for treating a pathology
CC mediated by TNFalpha, such as rheumatoid or osteo-arthritis. TNFalpha
CC mediated diseases which can be treated by the antibody include sepsis,
CC congestive heart failure, septic or endotoxic shock, cachexia, adult
CC respiratory distress syndrome, acquired immunodeficiency syndrome (AIDS),
CC allergies, psoriasis, tuberculosis, inflammatory bone disorders, blood
CC coagulation disorders, burns, rejection episodes following organ or
CC tissue transplant, Crohn's disease and autoimmune diseases, such as
CC thyroiditis. The antibodies may also be used to reduce the side effects
CC associated with TNFalpha generation during neoplastic therapy, to
CC eliminate or reduce shock-related symptoms associated with the treatment
CC or prevention of graft rejection by use of an anti-lymphocyte antibody,
CC for treating multi-organ failure, or in the diagnosis and imaging of
CC disease states involving elevated levels of TNF alpha. The present
CC sequence is hybrid complementarity determining region H2' (CDRH2')
CC peptide of human tumour necrosis factor (TNF) 40 heavy chain variable
CC domain.
XX
XX Sequence 17 AA;
SQ

Query Match 100.0%; Score 27; DB 23; Length 17;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 8 EPIYA 12

RESULT 2
ABB42462
ID ABB42462 standard; Peptide; 47 AA.
XX
AC ABB42462;
XX

04-FEB-2002 (first entry)
XX
XX Peptide #9968 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX

30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX

(MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
DR

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
XX
XX Claim 27; SEQ ID NO 35097; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 47 AA;
SQ

Query Match 100.0%; Score 27; DB 22; Length 47;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 16 EPIYA 20

RESULT 3
ABB25897
ID ABB25897 standard; Protein; 47 AA.
XX
AC ABB25897;
XX

23-JAN-2002 (first entry)
XX
XX Protein #7896 encoded by probe for measuring heart cell gene expression.
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX Homo sapiens.
XX WO200157274-A2.
XX

09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX

Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX

Single exon nucleic acid probes for analyzing gene expression in human hearts -

Claim 15; SEQ ID NO 27667; 530pp; English.

The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting, the
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 47 AA;

Query Match 100.0%; Score 27; DB 22; Length 47;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
| | | | |
Db 16 EPIYA 20

RESULT 4

AM63351
ID AAM63351 standard; Protein; 47 AA.

XX AC AAM63351;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 35456.

XX KW Human: brain expressed exon; gene expression analysis; probe;

XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer.

XX OS Homo. sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00667.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human
brains -

XX PS Example 4; SEQ ID NO: 35456; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.

XX SQ Sequence 47 AA;

Query Match 100.0%; Score 27; DB 22; Length 47;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
| | | | |
Db 16 EPIYA 20

RESULT 5

AAM76165
ID AAM76165 standard; Protein; 47 AA.

XX AC AAM76165;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 36471.
XX KW Human: bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00668.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488900/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human bone marrow -

XX PS Example 4; SEQ ID NO: 36471; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.

XX SQ Sequence 47 AA;

Query Match 100.0%; Score 27; DB 22; Length 47;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
| | | | |
Db 16 EPIYA 20

RESULT 6

AM20819
ID AAM20819 standard; Protein; 47 AA.

XX AC AAM20819;

XX DT 12-OCT-2001 (first entry)

XX PE Peptide #7253 encoded by probe for measuring cervical gene expression.

```
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW KW cervical cancer.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00670.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX PS Claim 27; SEQ ID No 25645; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX CC can be used to produce a single exon microarray, which can be used for
XX CC measuring human gene expression in a sample derived from human cervical
XX CC epithelial cells. By measuring gene expression, the probes are therefore
XX CC useful in grading and/or staging of diseases of the cervix, notably
XX CC cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX AC Sequence 47 AA;
XX SQ
XX Query Match 100.0%; Score 27; DB 22; Length 47;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 EPIYA 5
XX Db 16 EPIYA 20
XX RESULT 7
XX AAM36275
XX ID AAM36275 standard; Protein; 47 AA.
XX AC AAM36275;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #10312 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-234687P.
XX PR 04-OCT-2000; 2000US-236359P.
```

```
PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488997/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 27; SEQ ID No 36544; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX SQ Sequence 47 AA;
XX Query Match 100.0%; Score 27; DB 22; Length 47;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 EPIYA 5
XX Db 16 EPIYA 20
XX RESULT 8
XX ABG45519
XX ID ABG45519 standard; Peptide; 47 AA.
XX AC ABG45519;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 35184.
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW Chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-234687P.
XX PR 27-SEP-2000; 2000US-236359P.
```

PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX Claim 27; SEQ ID No 35184; 534pp; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequences by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 47 AA;
Query Match 100.0%; Score 27; DB 23; Length 47;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EPIYA 5
Db 16 EPIYA 20
RESULT 9
AAW23007
ID AAW23007 standard; Protein; 85 AA.
XX AC AAW23007;
XX 20-FEB-1998 (first entry)
XX

DE Canine herpesvirus unique long region PCUL49-85.
XX Vaccine; vector; gene therapy; canid; dog; CHV; CgE; PCUL49-85;
KW unique long region 49.
XX Canine herpesvirus.
OS WO9729772-A1.
PN 21-AUG-1997.
XX 14-FEB-1997; 97WO-US04115.
PF 15-FEB-1996; 96US-0602010.
PR (HESK-) HESKA CORP.
XX Frank RA, Haanes EJ;
PI WPI; 1997-424758/39.
XX N-PSDB; AAT75633.
DR Recombinant canine herpes virus and its genome - useful as vaccine
XX to protect canids against infectious, metabolic or genetic diseases
PT Claim 57; Page 210; 240pp; English.
XX This protein comprises canine herpesvirus (CHV) unique long region
PS 49 (PCUL49-85). Its sequence was deduced from a coding region
XX found in CHV genomic DNA molecule NCUL1823 (see AAT75633).
CC PCUL49-85 can be expressed in transformed host cells. CHV
CC proteins, nucleic acids, and antibodies raised against CHV
CC proteins, can be used to protect canids against CHV infection.
CC Novel recombinant CHV and novel recombinant CHV genomes that
CC contain heterologous nucleic acid molecules, can be used as
CC vaccines to protect canids against infectious, metabolic or genetic
CC diseases.
XX SQ Sequence 85 AA;
Query Match 100.0%; Score 27; DB 18; Length 85;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EPIYA 5
Db 12 EPIYA 16
RESULT 10
AAW72669
ID AAW72669 standard; Protein; 85 AA.
XX AC AAW72669;
XX 07-JAN-1999 (first entry)
XX Canine herpes virus protein sequence PCUL49-85.
DE Canine herpes virus; CHV; recombinant canine herpes virus vector;
XX genome; vaccination; dog; protozoan; helminth; ectoparasite; bacteria;
KW virus infection.
XX Canine herpes virus.
OS US5804197-A.
PN 08-SEP-1998.
PD 12-JUL-1996; 96US-0680726.
XX 12-JUL-1996; 96US-0680726.
PR 15-FEB-1996; 96US-0602010.

XX The present invention describes an isolated canine herpes virus (CHV)
CC (I) CDPase protein encoded by a CHV nucleic acid molecule that
CC hybridises under stringent hybridisation conditions with the CDPase
CC gene. Also described is a therapeutic composition (II) comprising (I).
CC (I) has antiallergic, immunosuppressive, cytostatic, antibacterial and
CC antiinflammatory activities, and can be used in vaccines. (I) is useful
CC for protecting an animal, in particular a canid from CHV. Therapeutic
CC compositions comprising (I) are useful for treating infectious diseases,
CC genetic diseases and other metabolic diseases, including diseases that
CC lead to abnormal cell growth, degenerative processes and/or
CC immunological defects, including allergies, autoimmune diseases,
CC cancers, cardiovascular diseases, graft rejection, haematopoietic
CC disorders, immunodeficiency diseases, immunoproliferative diseases,
CC immunosuppressive disorders, inflammatory diseases, jaundice, and septic
CC shock. CHV need not be attenuated for use as a live vaccine vector due
CC to the low pathogenicity of natural CHV, particularly as compared to
CC that of other herpesviruses. AAF26720 to AAF26786 and AAB51303 to
CC AAB51327 represent CHV nucleotides, proteins and PCR primers used in the
CC exemplification of the present invention.

XX SQ Sequence 85 AA;

Query Match 100.0%; Score 27; DB 22; Length 85;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 12 EPIYA 16
|||||

RESULT 12
ID AAG92796
AC AAG92796; Protein: 93 AA.
XX AAG92796;
XX 26-SEP-2001 (first entry)
XX C glutamicum protein fragment SEQ ID NO: 6550.
DE
XX Coryneform bacterium; amino acid synthesis: vitamin; saccharide;
KW organic acid synthesis.
KW Corynebacterium glutamicum.
OS
XX EPI108790-A2.
PN
XX 20-JUN-2001.
PD
XX 18-DEC-2000; 2000EP-0127688.
PF
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI
XX WPI; 2001-376931/40.
DR N-PSDB; AAH68015.
DR
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 6550; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 93 AA;
 Query Match 100.0%; Score 27; DB 22; Length 93;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EPIYA 5
 |||||
 Db 62 EPIYA 66

RESULT 13

AY87250
 ID AAY87250 standard; Protein: 115 AA.

XX AC AAY87250;

XX DT 11-MAY-2000 (first entry)

XX DE Human signal peptide containing protein HSPP-27 SEQ ID NO:27.

XX Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
 KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
 KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
 KW reproductive disorder; developmental disorder; arteriosclerosis;
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;
 KW muscular dystrophy.

XX OS Homo sapiens.

XX PN WO200000610-A2.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99WO-US14484.

XX PR 26-JUN-1998; 98US-0090762.

XX PR 31-JUL-1998; 98US-0094983.

XX PR 01-OCT-1998; 98US-0102686.

XX PR 11-DEC-1998; 98US-0112129.

XX (INCY-) INCYTE PHARM INC.

XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;

XX Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;

XX Bandman O;

XX WPI; 2000-160673/14.

XX N-PSDB; AAZ98135.

XX New human signal peptide-containing proteins useful in treatment,

XX prevention and diagnosis of e.g. cancer, inflammation and

XX cardiovascular disease -

XX Claim 1; Page 177; 327pp; English.

XX AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the

CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
 CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can
 CC be used in gene therapy. HSPPs can be used to treat or prevent disorders
 CC associated with decreased activity or function of HSPP. Antagonists of
 CC HSPP are used to treat or prevent disorders associated with increased
 CC activity or function of HSPP. Such diseases include cell proliferation
 CC (including cancer), inflammation, cardiovascular, neurological,
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
 CC asthma, Crohn's disease, microbial or other infections, congestive or
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
 CC nucleic acids can be used for the recombinant production of HSPP, for
 CC detecting HSPP in standard hybridisation and amplification assays (for
 CC diagnosis and monitoring), in gene therapy, as antisense,
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences
 CC or genetic variations, and for chromosomal mapping. HSPP are also used to
 CC raise specific antibodies (Ab) and to screen for agonists and
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
 CC monitor, HSPP-related diseases (in usual immunoassays), as therapeutic
 CC antagonists, in competitive drug screens, and for purification of HSPP
 CC from natural sources.

XX SQ Sequence 115 AA;

Query Match 100.0%; Score 27; DB 21; Length 115;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EPIYA 5

|||||

Db 105 EPIYA 109

RESULT 14

ABB89722
 ID ABB89722 standard; Protein: 115 AA.

XX AC ABB89722;

XX DT 24-MAY-2002 (first entry)

XX DE Human polypeptide SEQ ID NO 2098.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.

XX PN WO200190304-A2.

XX PD 29-NOV-2001.

XX PF 18-MAY-2001; 2001WO-US16450.

XX PR 19-MAY-2000; 2000US-205515P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

XX DR N-PSDB; ABL90131.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and

XX prevention of neural, immune system, muscular, reproductive,

XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative

XX disorders -

PS Claim 11; SEQ ID NO 2098; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB9040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and antagonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

XX Sequence 115 AA;

Query Match 100.0%; Score 27; DB 23; Length 115;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
 |||||
 Db 105 EPIYA 109

RESULT 15
 AAEL19666
 ID AAEL19666 standard; Protein: 118 AA.

XX
 AC AAEL19666;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 DE Human ghlhTNF40.4 CDR grafted variable heavy chain (Vh) protein.
 XX
 KW Human; tumour necrosis factor 40; TNF40; osteopathic; cardiant; CDR; complementarity determining region; rheumatoid; osteo-arthritis; sepsis; congestive heart failure; shock; tissue transplant; tuberculosis; AIDS; Acquired Immune Deficiency Syndrome; adult respiratory distress syndrome; cachexia; allergy; psoriasis; blood coagulation disorder; thyroiditis; inflammatory bone disorder; Crohn's disease; autoimmune disease; burn; neoplasty therapy; immunomodulator; vulnery; graft rejection.
 KW
 XX Homo sapiens.
 OS
 XX
 PN WO200194585-A1.
 XX
 PD 13-DEC-2001.
 XX
 PF 05-JUN-2001; 2001WO-GB02477.
 XX
 PR 06-JUN-2000; 2000GB-0013810.
 XX
 PA (CELL-) CELTECH R & D LTD.
 XX
 PI Athwal DS, Brown DT, Weir ANC, Popplewell AG, Chapman AP, King DJ;
 XX
 DR WPI; 2002-216732/27.
 DR N-PSDB; AAD31281.
 XX
 PT New antibody specific for human tumor necrosis factor (TNF)-alpha,
 PT useful for treating TNF-alpha-mediated diseases, e.g. congestive heart
 PT failure, septic or endotoxic shock, cachexia, adult respiratory
 PT distress syndrome
 XX
 XX Example-1; Fig 10; 119pp; English.

XX The invention relates to an antibody molecule having specificity for human tumour necrosis factor-alpha (TNFalpha) comprising a heavy or light chain. The antibody or the compound comprising the antibody is useful for treating or manufacturing a medicament for treating a pathology mediated by TNFalpha, such as rheumatoid or osteo-arthritis. TNFalpha mediated diseases which can be treated by the antibody include sepsis, congestive heart failure, septic or endotoxic shock, cachexia, adult respiratory distress syndrome, acquired immunodeficiency syndrome (AIDS), allergies, psoriasis, tuberculosis, inflammatory bone disorders, blood coagulation disorders, burns, rejection episodes following organ or tissue transplant, Crohn's disease and autoimmune diseases, such as thyroiditis. The antibodies may also be used to reduce the side effects associated with TNFalpha generation during neoplasty therapy, to eliminate or reduce shock-related symptoms associated with the treatment or prevention of graft rejection by use of an anti-lymphocyte antibody, for treating multi-organ failure, or in the diagnosis and imaging of disease states involving elevated levels of TNF alpha. The present sequence is complementarity determining (CDR) grafted heavy chain human tumour necrosis factor (TNF) 40.4 (ghlTNF40.4) protein.

XX

XX Sequence 118 AA;

Query Match 100.0%; Score 27; DB 23; Length 118;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
 |||||
 Db 57 EPIYA 61

Search completed: May 12, 2003, 11:53:12
 Job time : 1.51282 secs


```

US-09-092-409-84
; Sequence 84, Application US/09092409
; Patent No. 6159478
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexann S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092.409
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/680.726
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-092-409-84

```

```

Query Match 100.0% Score 27; DB 4; Length 85;
Best Local Similarity 100.0% Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EPIYA 5
    |||||
Db 12 EPIYA 16

```

```

RESULT 3
US-08-975-762-24
; Sequence 24, Application US/08975762
; Patent No. 6207169
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

TREATMEN

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975.762
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Ehrlichia
US-08-975-762-24
; Query Match 100.0% Score 27; DB 4; Length 376;
; Best Local Similarity 100.0% Pred. No. 1.7e+02;
; Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EPIYA 5
    |||||
Db 333 EPIYA 337

```

```

RESULT 4
US-08-821-324-24
; Sequence 24, Application US/08821324
; Patent No. 6231869
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821.324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Ehrlichia

```

THERAP

US-08-821-324-24

Query Match 100.0%; Score 27; DB 4; Length 376;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

Db 333 EPIYA 337

RESULT 5

US-09-295-028-24

; Sequence 24, Application US/09295028
; Patent No. 6277381
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION
; FILE REFERENCE: 210121.439C4
; CURRENT APPLICATION NUMBER: US/09/295,028
; CURRENT FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Ehrlichia sp.
US-09-295-028-24

Query Match 100.0%; Score 27; DB 4; Length 376;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

Db 333 EPIYA 337

RESULT 6

US-09-106-582-24

; Sequence 24, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; THERAPY
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,582
; FILING DATE: 29-JUN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Ehrlichia
; US-09-106-582-24

Query Match 100.0%; Score 27; DB 4; Length 376;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

Db 333 EPIYA 337

RESULT 7

US-09-066-046-4

; Sequence 4, Application US/09066046A
; Patent No. 6204252
; GENERAL INFORMATION:
; APPLICANT: MURPHY, Cheryl
; STOREY, James
; BELTZ, Gerald A.
; COUGHLIN, Richard T.
; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTTIC
; EHRlichia AND METHODS OF USE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,046A
; FILING DATE: 24-Apr-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Superko, Colleen
; REGISTRATION NUMBER: 39,850
; REFERENCE/DOCKET NUMBER: 106.941.155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 748 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-066-046-4

Query Match 100.0%; Score 27; DB 4; Length 748;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||

Db 584 EPIYA 588

RESULT 8

US-08-470-260-5
; Sequence 5, Application US/08470260
; Patent No. 6077706

GENERAL INFORMATION:

; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino

; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful

; for Vaccines and Diagnostics

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street

; CITY: Emeryville

; STATE: California

; COUNTRY: USA

; ZIP: 94608-2916

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/470.260

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/256.848

; FILING DATE: 21-OCT-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: McClung, Barbara G.

; REGISTRATION NUMBER: 33,113

; REFERENCE/DOCKET NUMBER: 0316.001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (510) 601-2708

; TELEFAX: (510) 655-3542

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1147 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-470-260-5

Query Match 100.0%; Score 27; DB 3; Length 1147;

Best Local Similarity 100.0%; Pred. No. 5.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

|||||

Db 890 EPIYA 894

RESULT 9

US-08-471-491-5

; Sequence 5, Application US/08471491B

; Patent No. 6090611

; GENERAL INFORMATION:

; APPLICANT: Covacci, Antonello

; APPLICANT: Bugnoli, Massimo

; APPLICANT: Telford, John

; APPLICANT: Macchia, Giovanni

; APPLICANT: Rappuoli, Rino

; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And

; TITLE OF INVENTION: Diagnostics

; FILE REFERENCE: CHIR0044

; CURRENT APPLICATION NUMBER: US/08/471.491B

; CURRENT FILING DATE: 1995-06-06

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 1147

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-08-471-491-5

Query Match 100.0%; Score 27; DB 3; Length 1147;

Best Local Similarity 100.0%; Pred. No. 5.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

|||||

Db 890 EPIYA 894

RESULT 10

US-08-466-662-5

; Sequence 5, Application US/08466662B

; Patent No. 6130059

; GENERAL INFORMATION:

; APPLICANT: Covacci, Antonello

; APPLICANT: Bugnoli, Massimo

; APPLICANT: Telford, John

; APPLICANT: Macchia, Giovanni

; APPLICANT: Rappuoli, Rino

; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And

; Diagnostics

; FILE REFERENCE: CHIR0057

; CURRENT APPLICATION NUMBER: US/08/466.662B

; CURRENT FILING DATE: 1995-06-06

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 1147

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-08-466-662-5

Query Match 100.0%; Score 27; DB 4; Length 1147;

Best Local Similarity 100.0%; Pred. No. 5.1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

|||||

Db 890 EPIYA 894

RESULT 11

US-08-053-614-4

; Sequence 4, Application US/08053614

; Patent No. 5403924

; GENERAL INFORMATION:

; APPLICANT: COVER, TIMOTHY L.

; APPLICANT: BLASER, MARTIN J.

; APPLICANT: TUMURU, MURALI K.R.

; TITLE OF INVENTION: THE tagA GENE AND METHODS FOR DETECTING

; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NEEDLE & ROSENBERG, P.C.

; STREET: 133 Carnegie Way, Suite 400

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: USA

; ZIP: 30303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/053,614
; FILING DATE: 19930426
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-053-614-4

Query Match 100.0%; Score 27; DB 1; Length 1181;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EPIYA 5
Db 891 EPIYA 895

RESULT 12
US-08-316-397B-4
; Sequence 4, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; TITLE OF INVENTION: THE taga GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-316-397B-4

Query Match 100.0%; Score 27; DB 1; Length 1181;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EPIYA 5
Db 891 EPIYA 895

Query Match 100.0%; Score 27; DB 1; Length 1181;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EPIYA 5
Db 891 EPIYA 895

RESULT 13
US-09-034-306-4
; Sequence 4, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; TITLE OF INVENTION: THE taga GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-4

Query Match 100.0%; Score 27; DB 2; Length 1181;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EPIYA 5
Db 891 EPIYA 895

RESULT 14
US-09-259-437-4
; Sequence 4, Application US/09259437
; Patent No. 6153390
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS

; APPLICANT: TUMMURU, MURALI K.R.
 ; TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
 ; AND PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 ; STREET: 127 Peachtree Street, Suite 1200
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/259,437
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/316,397
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: SPRATT, GWENDOLYN D.
 ; REGISTRATION NUMBER: 36,016
 ; REFERENCE/DOCKET NUMBER: 2200.030
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 404/688-0770
 ; TELEFAX: 404/688-9880
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1181 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-259-437-4

Query Match 100.0%; Score 27; DB 4; Length 1181;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
 Db 891 EPIYA 895

RESULT 15
 PCT-US93-09782-4
 ; Sequence 4. Application PC/TUS9309782
 ; GENERAL INFORMATION:
 ; APPLICANT: COVER, TIMOTHY L.
 ; APPLICANT: BLASER, MARTIN J.
 ; APPLICANT: TUMMURU, MURALI K.R.
 ; TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
 ; AND PREDISPOSITION TO PEPTIC ULCERATION
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 ; STREET: 133 Carnegie Way, Suite 400
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/09782
 ; FILING DATE:
 ; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:
 ; NAME: SPRATT, GWENDOLYN D.
 ; REGISTRATION NUMBER: 36,016
 ; REFERENCE/DOCKET NUMBER: 2200.009
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 404/688-0770
 ; TELEFAX: 404/688-9880
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1181 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US93-09782-4
 Query Match 100.0%; Score 27; DB 5; Length 1181;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
 Db 891 EPIYA 895

Search completed: May 12, 2003, 11:57:21
 Job time : 1.18376 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:55:34 ; Search time 0.247863 Seconds
(without alignments)
1856.381 Million cell updates/sec

Title: US-09-360-685B-10

Perfect score: 27 EPIYA 5

Sequence: 1 EPIYA 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US06_PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US07_PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	17	9	US-09-875-221A-2
2	27	100.0	17	10	US-09-949-559-2
3	27	100.0	17	10	US-09-864-761-41195
4	27	100.0	85	9	US-10-156-275-84
5	27	100.0	93	9	US-09-738-626-6550
6	27	100.0	118	9	US-09-875-221A-13
7	27	100.0	118	9	US-09-875-221A-15
8	27	100.0	118	10	US-09-949-559-13
9	27	100.0	118	10	US-09-949-559-15
10	27	100.0	150	9	US-10-001-631-2
11	27	100.0	216	9	US-09-875-221A-126
12	27	100.0	216	9	US-09-949-559-126
13	27	100.0	229	9	US-09-875-221A-130
14	27	100.0	229	10	US-09-949-559-130
15	27	100.0	244	9	US-09-880-748-1513
16	27	100.0	376	10	US-09-159-469-24
17	27	100.0	376	10	US-09-798-042-24
18	27	100.0	506	9	US-10-108-605-295
19	27	100.0	712	10	US-09-798-042-92

Sequence 58, Appl
Sequence 87, Appl
Sequence 4, Appl
Sequence 390, Appl
Sequence 387, Appl
Sequence 35675, A
Sequence 389, Appl
Sequence 39224, A
Sequence 5, Appl
Sequence 21, Appl
Sequence 156, Appl
Sequence 95, Appl
Sequence 13, Appl
Sequence 4, Appl
Sequence 723, Appl
Sequence 8, Appl
Sequence 69, Appl
Sequence 443, Appl
Sequence 62, Appl
Sequence 2, Appl
Sequence 10954, A
Sequence 3, Appl
Sequence 10, Appl
Sequence 4377, Appl
Sequence 7, Appl
Sequence 11, Appl

20 27 100.0 775 9 US-09-820-843A-58
21 27 100.0 1054 10 US-09-798-042-87
22 27 100.0 1338 10 US-09-402-100-4
23 26 96.3 30 9 US-09-983-802-390
24 26 96.3 38 9 US-09-983-802-387
25 26 96.3 59 10 US-09-864-761-35675
26 26 96.3 82 9 US-09-983-802-389
27 26 96.3 136 10 US-09-864-761-39224
28 26 96.3 494 9 US-09-981-947A-5
29 26 96.3 551 10 US-09-818-143-21
30 26 96.3 1374 9 US-10-121-988-158
31 26 96.3 1381 9 US-10-121-988-95
32 26 96.3 1436 9 US-10-080-505-13
33 25 92.6 235 9 US-10-007-706-4
34 25 92.6 372 10 US-09-925-297-723
35 25 92.6 499 9 US-09-994-064-8
36 25 92.6 499 9 US-09-994-064-69
37 25 92.6 883 10 US-09-741-669-443
38 24 88.9 114 9 US-09-895-913A-62
39 24 88.9 262 10 US-10-260-877-2
40 24 88.9 262 10 US-09-815-242-10954
41 24 88.9 383 9 US-10-285-074-3
42 24 88.9 493 12 US-10-002-593-10
43 24 88.9 985 9 US-09-738-626-4377
44 24 88.9 1091 9 US-09-925-388-7
45 23 85.2 12 10 US-09-966-871-11

ALIGNMENTS

RESULT 1
US-09-875-221A-2
; Sequence 2, Application US/09875221A
; Publication No. US20030026805A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products
; FILE REFERENCE: Carp-0089
; CURRENT APPLICATION NUMBER: US/09/875, 221A
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: GB0013810.7
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hTNF40/human hybrid CDRH2
US-09-875-221A-2

Query Match 100.0%; Score 27; DB 9; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 8 EPIYA 12

RESULT 2
US-09-949-559-2
; Sequence 2, Application US/09949559
; Patent No. US20020151682A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh

APPLICANT: Brown, Derek Thomas
 APPLICANT: Weir, Andrew Neil Charles
 APPLICANT: Popplewell, Andrew George
 APPLICANT: Chapman, Andrew Paul
 APPLICANT: King, David John
 TITLE OF INVENTION: Biological Products
 FILE REFERENCE: Carp-0095
 CURRENT APPLICATION NUMBER: US/09/949,559
 PRIOR FILING DATE: 2001-12-20
 PRIOR APPLICATION NUMBER: 0013810.7GB
 PRIOR FILING DATE: 2000-06-06
 PRIOR APPLICATION NUMBER: 09/875,221
 PRIOR FILING DATE: 2001-06-06
 NUMBER OF SEQ ID NOS: 130
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 2
 LENGTH: 17
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: hTNF40/human hybrid CDRH2
 US-09-949-559-2

Query Match 100.0%; Score 27; DB 10; Length 17;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
 |||||
 DB 8 EPIYA 12

RESULT 3
 US-09-864-761-41195
 ; Sequence 41195, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aeonica-x-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 41195
 ; LENGTH: 47
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC007000.2
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.2
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
 ; OTHER INFORMATION: EST_HUMAN HIT: AW299353.1, EVALUATE 8.00e-11
 US-09-864-761-41195

Query Match 100.0%; Score 27; DB 10; Length 47;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
 |||||
 DB 16 EPIYA 20

RESULT 4
 US-10-156-275-84
 ; Sequence 84, Application US/10156275
 ; Publication No. US20030049844A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haenes, Elizabeth J.
 ; APPLICANT: Frank, Rexann S.
 ; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
 ; NUMBER OF SEQUENCES: 92
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sheridan Ross & McIntosh
 ; STREET: 1700 Lincoln Street, Suite 3500
 ; CITY: Denver
 ; STATE: Colorado
 ; COUNTRY: U.S.A.
 ; ZIP: 80203
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/156,275
 ; FILING DATE: 28-May-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/399,118
 ; FILING DATE: <Unknown>
 ; APPLICATION NUMBER: 08/680,726
 ; FILING DATE: 12-JUL-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Connell, Gary J.
 ; REGISTRATION NUMBER: 32,020
 ; REFERENCE/DOCKET NUMBER: 2618-46-C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (303) 863-9700

; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-156-275-84

Query Match 100.0%; Score 27; DB 9; Length 85;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|
Db 12 EPIYA 16

RESULT 5

US-09-738-626-6550
; Sequence 6550, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6550
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6550

Query Match 100.0%; Score 27; DB 9; Length 93;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|
Db 62 EPIYA 66

RESULT 6

US-09-875-221A-13
; Sequence 13, Application US/09875221A
; Publication No. US20030026805A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products

; FILE REFERENCE: Carp-0089
; CURRENT APPLICATION NUMBER: US/09/875, 221A
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: GB0013810.7
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: gh1hTNF40.4 (Figure 10)
US-09-875-221A-13

Query Match 100.0%; Score 27; DB 9; Length 118;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|
Db 57 EPIYA 61

RESULT 7

US-09-875-221A-15
; Sequence 15, Application US/09875221A
; Publication No. US20030026805A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products
; FILE REFERENCE: Carp-0089
; CURRENT APPLICATION NUMBER: US/09/875, 221A
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: GB0013810.7
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: gh3hTNF40.4 (Figure 11)
US-09-875-221A-15

Query Match 100.0%; Score 27; DB 9; Length 118;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|
Db 57 EPIYA 61

RESULT 8

US-09-949-559-13
; Sequence 13, Application US/09949559
; Patent No. US20020151682A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products
; FILE REFERENCE: Carp-0095

; CURRENT APPLICATION NUMBER: US/09/949,559
; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 0013810.7GB
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 09/875,221
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: gh3HTNF40.4 (Figure 10)
US-09-949-559-13

Query Match 100.0%; Score 27; DB 10; Length 118;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||
Db 57 EPIYA 61

RESULT 9
US-09-949-559-15
; Sequence 15, Application US/09949559
; Patent No. US20020151682A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products
; FILE REFERENCE: Carp-0095
; CURRENT APPLICATION NUMBER: US/09/949,559
; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 0013810.7GB
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 09/875,221
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 15
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: gh3HTNF40.4 (Figure 11)
US-09-949-559-15

Query Match 100.0%; Score 27; DB 10; Length 118;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||
Db 57 EPIYA 61

RESULT 10
US-10-001-631-2
; Sequence 2, Application US/10001631
; Patent No. US20020164701A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Jellinek, Laura J.
; APPLICANT: Whitmore, Theodore E.
; TITLE OF INVENTION: Human Gene Marker for Metabolic Disease
; FILE REFERENCE: 98-16C1

; CURRENT APPLICATION NUMBER: US/10/001,631
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/105,450
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: 60/141,519
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 09/422,052
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-001-631-2

Query Match 100.0%; Score 27; DB 9; Length 150;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||
Db 140 EPIYA 144

RESULT 11
US-09-875-221A-126
; Sequence 126, Application US/09875221A
; Publication No. US20030026805A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products
; FILE REFERENCE: Carp-0089
; CURRENT APPLICATION NUMBER: US/09/875,221A
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: GB0013810.7
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 126
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Grafted heavy chain for fab
US-09-875-221A-126

Query Match 100.0%; Score 27; DB 9; Length 216;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||
Db 57 EPIYA 61

RESULT 12
US-09-949-559-126
; Sequence 126, Application US/09949559
; Patent No. US20020151682A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products

; FILE REFERENCE: Carp-0095
; CURRENT APPLICATION NUMBER: US/09/949,559
; CURRENT FILING DATE: 2001-12-20
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 0013810.7GB
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 09/875,221
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 126
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Grafted heavy chain for fab
US-09-949-559-126

Query Match 100.0%; Score 27; DB 10; Length 216;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EPIYA 5
Db 57 EPIYA 61
|||||

RESULT 13
US-09-875-221A-130
; Sequence 130, Application US/09875221A
; Publication No. US20030026805A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John
; TITLE OF INVENTION: Biological Products
; FILE REFERENCE: Carp-0089
; CURRENT APPLICATION NUMBER: US/09/875,221A
; CURRENT FILING DATE: 2001-06-06
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: GB0013810.7
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 130
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Grafted heavy chain for modified fab
US-09-875-221A-130

Query Match 100.0%; Score 27; DB 9; Length 229;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EPIYA 5
Db 57 EPIYA 61
|||||

RESULT 14
US-09-949-559-130
; Sequence 130, Application US/09949559
; Patent No. US20020151682A1
; GENERAL INFORMATION:
; APPLICANT: Athwal, Diljeet Singh
; APPLICANT: Brown, Derek Thomas
; APPLICANT: Weir, Andrew Neil Charles
; APPLICANT: Popplewell, Andrew George
; APPLICANT: Chapman, Andrew Paul
; APPLICANT: King, David John

; TITLE OF INVENTION: Biological Products
; FILE REFERENCE: Carp-0095
; CURRENT APPLICATION NUMBER: US/09/949,559
; CURRENT FILING DATE: 2001-12-20
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 0013810.7GB
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 09/875,221
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 130
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Grafted heavy chain for modified fab
US-09-949-559-130

Query Match 100.0%; Score 27; DB 10; Length 229;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EPIYA 5
Db 57 EPIYA 61
|||||

RESULT 15
US-09-880-748-1513
; Sequence 1513, Application US/09880748
; Publication No. US2003005937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1513
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1513

Query Match 100.0%; Score 27; DB 9; Length 244;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 EPIYA 5
Db 57 EPIYA 61
|||||

Search completed: May 12, 2003, 12:13:36
Job time : 1.24786 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:49:53 ; Search time 0.222222 Seconds
(without alignments)
2163.024 Million cell updates/sec

Title: US-09-360-685B-10

Perfect score: 27

Sequence: 1 EPIYA 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	27	100.0	80	2 B84285	hypothetical prote
2	27	100.0	305	2 T16567	hypothetical prote
3	27	100.0	324	2 H72637	hypothetical prote
4	27	100.0	372	2 I64223	bifunctional endo-
5	27	100.0	421	2 E71363	hypothetical prote
6	27	100.0	433	2 S26846	transcription fact
7	27	100.0	439	2 S29291	transcription fact
8	27	100.0	474	2 T16413	hypothetical prote
9	27	100.0	493	2 A41751	glucose transport
10	27	100.0	493	2 S38981	glucose transport
11	27	100.0	496	2 A31986	glucose transport
12	27	100.0	506	1 CSFF	catalase (EC 1.11.
13	27	100.0	506	2 T46196	cytochrome P450-11
14	27	100.0	512	2 S37279	hexon protein - hu
15	27	100.0	748	2 T08612	ankyrin-like prote
16	27	100.0	775	2 B72074	hypothetical prote
17	27	100.0	775	2 C81594	hypothetical prote
18	27	100.0	775	2 D86349	hypothetical prote
19	27	100.0	1096	2 C87263	hypothetical prote
20	27	100.0	1167	2 B71924	cag island protein
21	27	100.0	1186	2 C64588	cag pathogenicity
22	27	100.0	1215	2 B48281	cytotoxin-associat
23	27	100.0	1291	2 H71143	probable ribonucle
24	27	100.0	1842	2 T43409	probable fatty-aci
25	27	100.0	1842	2 T38781	fatty acid synthas
26	26	96.3	190	2 T33095	hypothetical prote
27	26	96.3	254	2 C75219	probable myo-inosi
28	26	96.3	277	2 S47835	hypothetical 30.7K
29	26	96.3	277	2 D91190	hypothetical prote

30	26	96.3	277	2 E86037	hypothetical prote
31	26	96.3	277	2 T45332	hypothetical prote
32	26	96.3	320	2 H95233	L-asparaginase, pr
33	26	96.3	320	2 B98098	asparaginase (EC 3
34	26	96.3	376	2 AG2227	DNA repair and gen
35	26	96.3	379	2 E71296	probable spermidin
36	26	96.3	394	2 F69401	cell division prot
37	26	96.3	451	1 B70058	conserved hypotet
38	26	96.3	581	2 E69322	dolichol-p-glucose
39	26	96.3	630	1 F69073	NADP-reducing hydr
40	26	96.3	694	2 AG2353	translation elonga
41	26	96.3	779	2 S57805	aconitate hydratase
42	26	96.3	841	1 S24462	probable 3',5'-cyc
43	26	96.3	918	2 D88544	protein R08D7.6 [i
44	26	96.3	983	2 E86989	probable integral
45	26	96.3	988	2 T24668	hypothetical prote

ALIGNMENTS

RESULT 1

B84285

hypothetical protein vngl300h [imported] - Halobacterium sp. NRC-1

C;Species: Halobacterium sp. NRC-1

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: B84285

R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A;Title: Genome sequence of Halobacterium species NRC-1.

A;Reference number: A84160; MUID:20504483; PMID:11016950

A;Accession: B84285

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-80 <STO>

A;Cross-references: GB:AE004437; MID:gl0580820; PIDN:AAG19646.1; GSPDB:GN00138

C;Genetics:

A;Gene: VNG1300H

Query Match 100.0%; Score 27; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

DB 47 EPIYA 51

RESULT 2

T16567

hypothetical protein K05F1.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C;Accession: T16567

R;Wohlmann, P.

submitted to the EMBL Data Library, June 1995

A;Description: The sequence of C. elegans cosmid K05F1.

A;Reference number: Z18537

A;Accession: T16567

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-305 <WOH>

A;Cross-references: EMBL:U29377; MID:g868173; PID:g868182; PIDN:AAA68719.1; CESP:K05F

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:K05F1.4

A;Introns: 44/2; 100/1; 164/1; 188/2; 295/1

Query Match 100.0%; Score 27; DB 2; Length 305;

Best Local Similarity 100.0%; Pred. No. 87;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||

Db 61 EPIYA 65

RESULT 3
H72637
hypothetical protein APE0537 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: H72637
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72637
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <KAW>
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79504.1; PID:g5104189
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0537
C:Superfamily: conserved hypothetical protein MJ1157

Query Match 100.0%; Score 27; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||

Db 302 EPIYA 306

RESULT 4
I64223
bifunctional endo-1,4-beta-xylanase homolog - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
C:Accession: I64223
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J. C.A.; Venter, J.C. Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: I64223
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-372 <TIGR>
A:Cross-references: GB:U39699; GB:L43967; NID:g1045903; PID:g1045904; TIGR:MG217
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3

Query Match 100.0%; Score 27; DB 2; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||

Db 355 EPIYA 359

RESULT 5
E71363
hypothetical protein TP0118 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: E71363

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M they, L.; Weldman, J.; Smith, H.O.; Venter, J.C. Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: E71363
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-421 <COL>
A:Cross-references: GB:AE001197; GB:AE000520; NID:g3322382; PIDN:AAC65114.1; PID:g3322
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0118

Query Match 100.0%; Score 27; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||

Db 123 EPIYA 127

RESULT 6
S26646
transcription factor IIE - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: S26646
R:Ohkuma, Y.; Hashimoto, S.; Roeder, R.G.; Horikoshi, M. Nucleic Acids Res. 20, 5838, 1992
A:Title: Identification of two large subdomains in TFIIIF-alpha on the basis of homolo
A:Reference number: S26646; MUID:93087200; PMID:1454543
A:Accession: S26646
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <OHK>
A:Cross-references: EMBL:Z14131; NID:g65130; PIDN:CAA78505.1; PID:g65131

Query Match 100.0%; Score 27; DB 2; Length 433;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
|||||

Db 185 EPIYA 189

RESULT 7
S29291
transcription factor IIE 56K chain - human
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
C:Accession: S29291; S29319
R:Peterson, M.G.; Inostroza, J.; Maxon, M.E.; Flores, O.; Admon, A.; Reinberg, D.; Tj Nature 354, 369-373, 1991
A:Title: Structure and functional properties of human general transcription factor II
A:Reference number: S29291; MUID:92065976; PMID:1956398
A:Accession: S29291
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-439 <PET>
A:Cross-references: GB:S67859; NID:g239577; PIDN:AAB20413.1; PID:g239578
R:Ohkuma, Y.; Sumimoto, H.; Hoffmann, A.; Shimasaki, S.; Horikoshi, M.; Roeder, R.G. Nature 354, 398-401, 1991
A:Title: Structural motifs and potential sigma homologues in the large subunit of hum
A:Reference number: S29319; MUID:92065982; PMID:1956403
A:Accession: S29319
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-351, 'D', 353-439 <OHK>
A:Cross-references: EMBL:X63468; NID:g37067; PIDN:CAA45068.1; PID:g37068

C;Keywords: transcription regulation

Query Match 100.0%; Score 27; DB 2; Length 439;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
 |||||
 Db 185 EPIYA 189

RESULT 8

TL6413

hypothetical protein F48E8.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000

C;Accession: T16413

R;Kirsten, J.

submitted to the EMBL Data Library, March 1995

A;Description: The sequence of C. elegans cosmid F48E8.

A;Reference number: S59413

A;Accession: T16413

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-474 <KIR>

A;Cross-references: EMBL:U23514; NID:g746484; PID:g746486; PIDN:AAC46539.1; CESP:F48E8.3

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:F48E8.3

A;Introns: 49/3; 238/1; 339/1; 380/3

C;Superfamily: osmotic growth protein 1; 3-oxosteroid 1-dehydrogenase homology; fumarate

F:10-281/Domain: fumarate reductase flavoprotein homology <FRF>

F:324-455/Domain: 3-oxosteroid 1-dehydrogenase homology <OXD>

Query Match 100.0%; Score 27; DB 2; Length 474;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
 |||||
 Db 391 EPIYA 395

RESULT 9

A41751

glucose transport protein 3 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992.#text_change 24-Sep-1999

C;Accession: A41751; S18128

R;Nagamatsu, S.; Kornhauser, J.M.; Burant, C.F.; Seino, S.; Mayo, K.E.; Bell, G.I.

J. Biol. Chem. 267, 467-472, 1992

A;Title: Glucose transporter expression in brain. cDNA sequence of mouse GLUT3, the brain

A;Reference number: A41751; MUID:92112695; PMID:1730609

A;Accession: A41751

A;Molecule type: mRNA

A;Residues: 1-493 <NAG>

A;Cross-references: GB:X61093; NID:g51088; PIDN:CAA43406.1; PID:g51089

C;Genetics:

A;Gene: GLUT3

C;Superfamily: glucose transport protein

C;Keywords: glycoprotein; transmembrane protein

F:43/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 27; DB 2; Length 493;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
 |||||
 Db 303 EPIYA 307

RESULT 10

S38981

glucose transport protein 3, neuron-specific - rat

N;Alternate names: glucose transporter

C;Species: Rattus norvegicus (Norway rat)

C;Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000

C;Accession: S38981

R;Nagamatsu, S.; Sawa, H.; Kamada, K.; Nakamichi, Y.; Yoshimoto, K.; Hoshino, T.

FEBS Lett. 334, 289-295, 1993

A;Title: Neuron-specific glucose transporter (NSGT): CNS distribution of GLUT3 rat gl

A;Reference number: S38981; MUID:94063043; PMID:8243635

A;Accession: S38981

A;Molecule type: mRNA

A;Residues: 1-493 <NAG>

A;Cross-references: EMBL:D13962; NID:g220748; PIDN:BAA03065.1; PID:g220749

A;Note: the authors translated the codon TAC for residue 141 as Thr, GTC for residue

C;Superfamily: glucose transport protein

C;Keywords: transmembrane protein

Query Match 100.0%; Score 27; DB 2; Length 493;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
 |||||
 Db 303 EPIYA 307

RESULT 11

A31986

glucose transporter-like protein, skeletal muscle - human

C;Species: Homo sapiens (man)

C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 24-Sep-1999

C;Accession: A31986

R;Kavano, T.; Fukumoto, H.; Eddy, R.L.; Fan, Y.S.; Byers, M.G.; Shows, T.B.; Bell, G.

J. Biol. Chem. 263, 15245-15248, 1988

A;Title: Evidence for a family of human glucose transporter-like proteins. Sequence a

A;Reference number: A31986; MUID:89008414; PMID:3170580

A;Accession: A31986

A;Molecule type: mRNA

A;Residues: 1-496 <KAY>

A;Cross-references: GB:M20681; NID:g183684; PIDN:AAB61083.1; PID:g306821

C;Genetics:

A;Gene: GDB:SLC2A3; GLUT3

A;Cross-references: GDB:I19996; OMIM:138170

A;Map position: 12p13.3-12p13.3

C;Superfamily: glucose transport protein

C;Keywords: muscle; skeletal muscle; transmembrane protein

Query Match 100.0%; Score 27; DB 2; Length 496;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
 |||||
 Db 303 EPIYA 307

RESULT 12

CSFF

catalase (EC 1.11.1.6) - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 03-Mar-2000

C;Accession: S12725; S69287; S15059

R;Orr, E.C.; Bewley, G.C.; Orr, W.C.

Nucleic Acids Res. 18, 3663, 1990

A;Title: cDNA and deduced amino acid sequence of Drosophila catalase.

A;Reference number: S12725; MUID:90301508; PMID:2362827

A;Accession: S12725

A;Molecule type: mRNA

A;Residues: 1-506 <ORR1>

A;Cross-references: EMBL:X52286; NID:g7689; PIDN:CAA36529.1; PID:g7690

R;Orr, W.C.; Orr, E.C.; Legan, S.K.; Sohal, R.S.

Arch. Biochem. Biophys. 330, 251-258, 1996
 A:Title: Molecular analysis of the Drosophila catalase gene.
 A:Reference number: S69287; MUID:96239139; PMID:8660653
 A:Accession: S69287
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-506 <ORF>
 A:Cross-references: EMBL:U00145; NID:g451307; PIDN:AAC13738.1; PID:g451308

C:Genetics:
 A:Gene: FlyBase:Cat
 A:Cross-references: FlyBase:FBgn0000261
 A:Map position: 3L 75D-E1
 A:Introns: 19/3; 326/3
 C:Superfamily: catalase
 C:Keywords: Chromoprotein; heme; iron; metalloprotein; oxidoreductase
 F:73,112,146/Active site: His, Ser, Asn #status predicted
 F:356/Binding site: heme iron (Tyr) (axial ligand) #status predicted

Query Match 100.0%; Score 27; DB 1; Length 506;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
 |||||

Db 226 EPIYA 230

RESULT 13

T46196
 cytochrome P450-like protein - Arabidopsis thaliana
 N:Alternate names: protein T8P19.30
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001
 C:Accession: T46196
 R:Cholsne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223008

A:Accession: T46196

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 <CHO>

A:Cross-references: EMBL:AL133315

A:Experimental source: cultivar Columbia; BAC clone T8P19

C:Genetics:

A:Map position: 3

A:Note: T8P19.30

C:Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology

F:300-469/Domain: cytochrome P450 homology <P45>

Query Match 100.0%; Score 27; DB 2; Length 506;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

|||||

Db 225 EPIYA 229

RESULT 14

S37279
 hexon protein - human adenovirus 9 (fragment)
 C:Species: Mastadenovirus h9 (human adenovirus 9)
 C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 26-Aug-1999
 C:Accession: S37279; S37221

R:Eliz, B.

submitted to the EMBL Data Library, August 1993

A:Reference number: S37276

A:Accession: S37279

A:Molecule type: DNA

A:Residues: 1-512 <EIZ>

A:Cross-references: EMBL:X74657; NID:g404270; PIDN:CAA52721.1; PID:g404271

R:Pring-Akerblom, P.

submitted to the EMBL Data Library, September 1993

A:Reference number: S37213
 A:Accession: S37221
 A:Molecule type: DNA
 A:Residues: 1-512 <PRI>
 A:Cross-references: EMBL:X74664; NID:g402773; PIDN:CAA52728.1; PID:g402774
 C:Superfamily: adenovirus hexon protein
 C:Keywords: hexon protein

Query Match 100.0%; Score 27; DB 2; Length 512;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

|||||

Db 79 EPIYA 83

RESULT 15

T08612
 ankyrin-like protein, 160K - Ehrlichia sp. (strain USG3)
 C:Species: Ehrlichia sp.
 A:Variety: strain USG3
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T08612
 R:Storey, J.R.; Doros-Richert, L.A.; Gingrich-Baker, C.; Munroe, K.; Mather, T.N.; Co
 Infect. Immun. 66, 1356-1363, 1998
 A:Title: Molecular cloning and sequencing of three granulocytic Ehrlichia genes encod
 A:Reference number: Z16457; MUID:98187902; PMID:9529053
 A:Accession: T08612
 A>Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-748 <STO>

A:Cross-references: EMBL:AF020521; NID:g3033493; PID:g3033494

A:Experimental source: strain USG3

C:Genetics:

A:Gene: ank

Query Match 100.0%; Score 27; DB 2; Length 748;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

|||||

Db 584 EPIYA 588

Search completed: May 12, 2003, 11:56:37
 Job time : 2.22222 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:42:53 ; Search time 0.132479 Seconds
(without alignments)
1565.397 Million cell updates/sec

Title: US-09-360-685B-10

Perfect score: 27

Sequence: 1 EPIYA 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	27	100.0	171	1	FABA_RHILO
2	27	100.0	372	1	P65H_MYCGE
3	27	100.0	400	1	GTR3_RABIT
4	27	100.0	421	1	Y118_TREPA
5	27	100.0	439	1	T2EA_HUMAN
6	27	100.0	440	1	T2EA_MOUSE
7	27	100.0	493	1	GTR3_MOUSE
8	27	100.0	493	1	GTR3_RAT
9	27	100.0	495	1	GTR3_CANFA
10	27	100.0	496	1	GTR3_HUMAN
11	27	100.0	506	1	CATA_DROME
12	27	100.0	512	1	HEX_ADE09
13	27	100.0	786	1	APEL_SULTO
14	27	100.0	1147	1	CGA1_HELPY
15	27	100.0	1167	1	CAGA_HELPY
16	27	100.0	1182	1	CGA2_HELPY
17	27	100.0	1186	1	CAGA_HELPY
18	27	100.0	1842	1	FAS2_SCHPO
19	26	96.3	319	1	YIBQ_ECOLI
20	26	96.3	376	1	RECF_ANASP
21	26	96.3	424	1	LE21_METAC
22	26	96.3	443	1	FUTB_DROME
23	26	96.3	494	1	GTR3_BOVIN
24	26	96.3	494	1	GTR3_SHEEP
25	26	96.3	779	1	ACON_GRAVE
26	26	96.3	918	1	YNE6_CABEL
27	26	96.3	983	1	Y644_MYCLE
28	26	96.3	992	1	YV93_MYCTU
29	26	96.3	1958	1	UBR1_SCHPO
30	25	92.6	233	1	Y308_METJA
31	25	92.6	260	1	YQ08_CABEL
32	25	92.6	302	1	TRUB_XYLFA
33	25	92.6	384	1	RECFC_SYNY3

RESULT 1

ID	FABA_RHILO	STANDARD	PRT	171 AA
AC	Q98BH8			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.60)			
DE	(beta-hydroxydecanoyl thioester dehydrase).			
GN	FABA OR MLL5569.			
OS	Rhizobium loti (Mesorhizobium loti).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Phyllobacteriaceae; Mesorhizobium.			
OX	NCBI_TaxID=381;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=MAFF303099;			
RX	MEDLINE=21082930; PubMed=11214968;			
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,			
RA	Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,			
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,			
RA	Takeuchi C., Yamada M., Tabata S.;			
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium			
RT	Mesorhizobium loti".			
RT	DNA Res. 7:331-338(2000).			
CC	-1- FUNCTION: Necessary for the introduction of cis unsaturation into			
CC	fatty acids. Catalyzed the dehydration of (3R)-3-hydroxydecanoyl-			
CC	ACP to E-(2)-decanoyl-ACP and then its isomerization to Z-(3)-			
CC	decanoyl-ACP (By similarity).			
CC	-1- CATALYTIC ACTIVITY: (3R)-3-hydroxydecanoyl-[acyl-carrier protein]			
CC	-> 2,3-decanoyl-[acyl-carrier protein] or 3,4-decanoyl-[acyl-			
CC	carrier protein] + H(2)O.			
CC	-1- PATHWAY: KEY STEP IN THE ANAEROBIC PATHWAY OF UNSATURATED FATTY			
CC	ACID SYNTHESIS IN BACTERIA.			
CC	-1- SUBUNIT: Homodimer (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE THIOESTER DEHYDRATASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL: AP003006; BAB51994.1;			
DR	InterPro: IPR001143; Bac_Sdehydratase.			
DR	Pfam: PF01377; Thioester_dehyd; 1.			
KW	Lysase; Fatty acid biosynthesis; Complete proteome.			
FT	ACT SITE 70 BY SIMILARITY.			
SQ	SEQUENCE 171 AA; 18706 MW; B6638EB79B9ADAEB CRC64;			

Query Match 100.0%; Score 27; DB 1; Length 171;
Best Local Similarity 100.0%; Pred. No. 14;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

Db 152 EPIYA 156

RESULT 2

P65H_MYCGE STANDARD; PRT; 372 AA.

AC P47459;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Proline-rich P65 protein homolog.

GN MG217.

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2097;

RN [1]

RC SEQUENCE FROM N.A.

RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,

RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,

RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,

RA Tomb J.F., Dougherty B.A., Boff K.F., Hu P.-C., Lucier T.S.,

RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

RT "The minimal gene complement of Mycoplasma genitalium";

RL Science 270:397-403(1995).

CC -1- SIMILARITY: TO M.PNEUMONIAE PROTEIN P65.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; U39701; AAC71436.1; -

DR TIGR; MG217; -

KW Repeat; Complete proteome.

FT DOMAIN 29 159 8 X DPNQ(Q)FNQ REPEATS.

FT REPEAT 29 40 1-1.

FT REPEAT 41 52 1-2.

FT REPEAT 53 60 2-1.

FT REPEAT 61 72 1-3.

FT REPEAT 73 80 2-2.

FT REPEAT 81 92 1-4.

FT REPEAT 93 100 2-3.

FT REPEAT 101 112 1-5.

FT REPEAT 113 119 2-4.

FT REPEAT 120 131 1-6.

FT REPEAT 132 138 2-5.

FT REPEAT 139 150 1-7.

FT REPEAT 151 162 1-8.

SQ SEQUENCE 372 AA; 44664 MW; 4C29701D213CE19E CRC64;

Query Match

Best Local Similarity 100.0%; Score 27; DB 1; Length 372;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

Db 355 EPIYA 359

RESULT 3

GTR3_RABIT

ID GTR3_RABIT

AC O9XSC2;

STANDARD;

PRT; 400 AA.

DT

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Solute carrier family 2, facilitated glucose transporter, member 3

DE (Glucose transporter type 3, brain) (Fragment).

GN SLC2A3 OR GLUT3.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=New Zealand white;

RX MEDLINE=99196804; PubMed=10095018;

RA Devaskar S.U., Rajakumar P.A., Mink R.B., McKnight R.A.,

RA Thamocharan S., Hicks S.M.;

RT "Effect of development and hypoxia-ischemia upon rabbit brain glucose

RT transporter expression";

RL Brain Res. 823:113-128(1999).

CC -1- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. PROBABLY A NEURONAL

CC GLUCOSE TRANSPORTER.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE

CC TRANSPORTER SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; AF117812; AAD26251.1; -

DR InterPro; IPR003663; CHO transporter.

DR InterPro; IPR003662; sub-transporter.

DR Pfam; PF00083; sugar_tr_1

DR PRINTS; PR00171; SUGRTNSPORT.

DR TIGRFAMS; TIGR00879; SP; 1.

DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.

DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.

KW Transmembrane; Sugar transport; Transp; Multigene family.

FT NON_TER 1 1

FT DOMAIN <1 4

FT TRANSMEM 5 25 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 26 27 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 28 48 4 (POTENTIAL).

FT DOMAIN 49 57 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 58 78 5 (POTENTIAL).

FT DOMAIN 79 84 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 85 105 6 (POTENTIAL).

FT DOMAIN 106 174 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 175 195 7 (POTENTIAL).

FT DOMAIN 196 209 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 210 230 8 (POTENTIAL).

FT DOMAIN 231 238 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 239 259 9 (POTENTIAL).

FT DOMAIN 260 273 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 274 294 10 (POTENTIAL).

FT DOMAIN 295 304 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 305 325 11 (POTENTIAL).

FT DOMAIN 326 332 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 333 353 12 (POTENTIAL).

FT DOMAIN 354 400 CYTOPLASMIC (POTENTIAL).

FT SITE 182 184 DEFINES SUBSTRATE SPECIFICITY (BY

FT SIMILARITY).

SQ SEQUENCE 400 AA; 43722 MW; F845D5E5248B9A79 CRC64;

Query Match

Best Local Similarity 100.0%; Score 27; DB 1; Length 400;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5

```

Db 208 EPIYA 212
|||||
RESULT 4
Y118_TREPA STANDARD; PRT; 421 AA.
AC 083155;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0118.
GN TP0118.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=96332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT *Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.*;
RL Science 281:375-388(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AEO01197; AAC65114.1; -
DR TIGR; TP0118; -
KW Hypothetical protein; Coiled coil; Complete proteome.
FT DOMAIN 126 182 COILED COIL (POTENTIAL).
FT DOMAIN 328 397 COILED COIL (POTENTIAL).
SQ SEQUENCE 421 AA; 46702 MW; 8C7C72BD0C515B6A CRC64;

Query Match 100.0%; Score 27; DB 1; Length 421;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 123 EPIYA 127
|||||
RESULT 5
T2EA_HUMAN STANDARD; PRT; 439 AA.
ID T2EA_HUMAN
AC P29083; Q16103;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor IIE, alpha subunit (TFIIE-alpha)
DE (General transcription factor IIE 56 kDa subunit).
GN GTF2E1 OR TP2E1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92065976; PubMed=1956398;
RA Peterson M.G., Inostroza J., Maxon M.E., Flores O., Admon A.,

Reinberg D., Tjian R.;
RT "Structure and functional properties of human general transcription
RT factor IIE.";
RL Nature 354:369-373(1991).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92065982; PubMed=1956403;
RA Ohkuma Y., Sumimoto H., Hoffmann A., Shimasaki S., Horikoshi M.,
RA Roeder R.G.;
RT "Structural motifs and potential sigma homologues in the large
RT subunit of human general transcription factor TFIIE.";
RL Nature 354:398-401(1991)
CC -1- FUNCTION: RECRUITS TFIID TO THE INITIATION COMPLEX AND STIMULATES
CC THE RNA POLYMERASE II C-TERMINAL DOMAIN KINASE AND DNA-DEPENDENT
CC ATPASE ACTIVITIES OF TFIID. BOTH TFIID AND TFIIE ARE REQUIRED FOR
CC PROMOTER CLEARANCE BY RNA POLYMERASE.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TFIIE ALPHA SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S67859; AAB20413.1; -
DR EMBL; X63468; CAA45068.1; -
DR PIR; S29319; S29319.
DR TRANSFAC; T02162; -
DR MIM; 189962; -
DR InterPro; IPR002853; TFIIE_alpha.
DR Pfam; PF02002; TFIIE_alpha.
DR SMART; SM00531; TFIIE; 1.
KW Transcription regulation; Nuclear protein; Zinc-finger.
FT ZN_FING 129 157 C4-TYPE (POTENTIAL).
FT DOMAIN 378 393 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 352 352 D -> S (IN REF. 1).
SQ SEQUENCE 439 AA; 49480 MW; 75228E84E737413 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 185 EPIYA 189
|||||
RESULT 6
T2EA_MOUSE STANDARD; PRT; 440 AA.
ID T2EA_MOUSE
AC Q9D0D5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor IIE, alpha subunit (TFIIE-alpha)
DE (General transcription factor IIE 56 kDa subunit).
GN GTF2E1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

```

RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavani T.,
RA Fieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayaishi Y.,
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -1- FUNCTION: RECRUITS TFIIF TO THE INITIATION COMPLEX AND STIMULATES
CC THE RNA POLYMERASE II C-TERMINAL DOMAIN KINASE AND DNA-DEPENDENT
CC APPASE ACTIVITIES OF TFIIF. BOTH TFIIF AND TFIIE ARE REQUIRED FOR
CC PROMOTER CLEARANCE BY RNA POLYMERASE.
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -1- SIMILARITY: BELONGS TO THE TFIIE ALPHA SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK011543; BAB27686.1; -
DR MGD; MGI:1921447; Gtf2el.
DR InterPro: IPR002853; TFIIE.alpha.
DR Pfam: PF02002; TFIIE.alpha.1.
DR SMART: SM00531; TFIIE; 1.
KW Transcription regulation; Nuclear protein; Zinc-finger.
FT ZN_FING 129 157 C4-TYPE (POTENTIAL).
FT DOMAIN 380 394 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 440 AA; 49593 MW; 0587E4DFE6FA0442 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 440;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
| | | | |
Db 185 EPIYA 189

RESULT 7
ID GTR3_MOUSE STANDARD; PRT; 493 AA.
AC P32037;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Solute carrier family 2, facilitated glucose transporter, member 3
DE (Glucose transporter type 3, brain).
GN SLC2A3 OR GLUT3 OR GLUT-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92112695; PubMed=1730609;
RA Nagamatsu S., Kornhauser J.M., Saino S., Mayo K.E., Steiner D.F.,
RA Bell G.I.;
RT "Glucose transporter expression in brain. cDNA sequence of mouse

RT GLUT3, the brain facilitative glucose transporter isoform, and
RT identification of sites of expression by in situ hybridization.";
RL J. Biol. Chem. 267:467-472(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Takeda J., Minokoshi Y., Yasuda K., Kayano T., Graeme B.I.;
RA Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. PROBABLY A NEURONAL
CC GLUCOSE TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC TRANSPORTER SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75135; AAA37704.1; -
DR EMBL; X61093; CAA43406.1; -
DR EMBL; U11853; AAB60666.1; -
DR EMBL; U11844; AAB60666.1; JOINED.
DR EMBL; U11845; AAB60666.1; JOINED.
DR EMBL; U11846; AAB60666.1; JOINED.
DR EMBL; U11848; AAB60666.1; JOINED.
DR EMBL; U11849; AAB60666.1; JOINED.
DR EMBL; U11850; AAB60666.1; JOINED.
DR EMBL; U11851; AAB60666.1; JOINED.
DR EMBL; U11852; AAB60666.1; JOINED.
DR PIR; A41751; A41751.
DR MGD; MGI:95757; Slc2a3.
DR InterPro: IPR003663; CHO transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_Cr; 1.
DR PRINTS; PR00171; SUGTRNSPORT.
DR TIGRFAMS; TIGR00879; SP; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane; Sugar transport; Transport; Glycoprotein;
KW Multigene family.
FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 11 31 1 (POTENTIAL).
FT DOMAIN 32 64 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 65 85 2 (POTENTIAL).
FT DOMAIN 86 93 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 94 114 3 (POTENTIAL).
FT DOMAIN 115 124 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 125 145 4 (POTENTIAL).
FT DOMAIN 146 153 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 154 174 5 (POTENTIAL).
FT DOMAIN 175 183 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 184 204 6 (POTENTIAL).
FT DOMAIN 205 269 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 270 290 7 (POTENTIAL).
FT DOMAIN 291 304 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 305 325 8 (POTENTIAL).
FT DOMAIN 326 334 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 335 355 9 (POTENTIAL).
FT DOMAIN 356 363 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 364 389 10 (POTENTIAL).
FT DOMAIN 385 399 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 400 420 11 (POTENTIAL).
FT DOMAIN 421 427 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 428 448 12 (POTENTIAL).
FT DOMAIN 449 493 CYTOPLASMIC (POTENTIAL).
FT SITE 277 DEFINES SUBSTRATE SPECIFICITY (BY
FT SIMILARITY).
FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 493 AA; 53478 MW; 9090B8DCB8780082 CRC64;

```

Query Match          100.0%; Score 27; DB 1; Length 493;
Best Local Similarity 100.0%; Pred. NO. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 EPIYA 5
DB      303 EPIYA 307

RESULT 8
GTR3 RAT
ID      GTR3_RAT      STANDARD;      PRT;      493 AA.
AC      Q07647; Q62729;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Solute carrier family 2, facilitated glucose transporter, member 3
DE      (Glucose transporter type 3, brain).
GN      SLC2A3 OR GLUT3 OR GLUT-3.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94033043; PubMed=8243635;
RA      Nagamatsu S., Sawa H., Kamada K., Nakamichi Y., Yoshimoto K.,
RA      Hoshino T.;
RT      "Neuron-specific glucose transporter (NSGT): CNS distribution of
RT      GLUT3 rat glucose transporter (RGT3) in rat central neurons.";
RL      PDBS Lett. 334:289-295(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=Sprague-Dawley; TISSUE=Brain;
RX      MEDLINE=96042338; PubMed=7475896;
RA      Krishnan S.N., Haddad G.G.;
RT      "Cloning of glucose transporter-3 (GLUT3) cDNA from rat brain.";
RL      Life Sci. 56:1193-1197(1995).
CC      -1- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. PROBABLY A NEURONAL
CC      GLUCOSE TRANSPORTER.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.
CC      -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC      TRANSPORTER SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@lsb-sib.ch).
-----
EMBL; D13962; BAA03065.1; --
EMBL; U17978; AAA62503.1; --
PIR; S38981; S38981.
InterPro; IPR003663; CHO transport.
InterPro; IPR003662; sub.transporter.
Pfam; PF00083; sugar_tr; 1.
PRINTS; PR00171; SUGRTNSPORT.
TIGRfams; TIGR00879; SP; 1.
PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
Transmembrane; Sugar transport; Transport; Glycoprotein;
Multigene family.
FT      DOMAIN          1 10  CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM        11 31  1 (POTENTIAL).
FT      DOMAIN          32 64  EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM        65 85  2 (POTENTIAL).
FT      DOMAIN          86 93  CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM        94 114 3 (POTENTIAL).
FT      DOMAIN          115 124 EXTRACELLULAR (POTENTIAL).

```

```

DR InterPro: IPR003663; CHO_transport.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR PRINTS: PR00171; SUGRTNSPORT.
DR TIGRFAMS: TIGR00879; SP; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane; Sugar transport; Transport; Glycoprotein;
KW Multigene family.
FT DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 6 26 1 (POTENTIAL).
FT DOMAIN 27 64 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 65 85 2 (POTENTIAL).
FT DOMAIN 86 93 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 94 114 3 (POTENTIAL).
FT DOMAIN 115 124 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 125 145 4 (POTENTIAL).
FT DOMAIN 146 153 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 154 174 5 (POTENTIAL).
FT DOMAIN 175 183 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 184 204 6 (POTENTIAL).
FT DOMAIN 205 269 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 270 290 7 (POTENTIAL).
FT DOMAIN 291 304 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 305 325 8 (POTENTIAL).
FT DOMAIN 326 334 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 335 355 9 (POTENTIAL).
FT DOMAIN 356 369 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 370 390 10 (POTENTIAL).
FT DOMAIN 391 399 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 400 420 11 (POTENTIAL).
FT DOMAIN 421 424 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 425 445 12 (POTENTIAL).
FT DOMAIN 446 495 CYTOPLASMIC (POTENTIAL).
FT SITE 277 279 DEFINES SUBSTRATE SPECIFICITY (BY SIMILARITY).
FT CARBOHYD 43 43 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 495 AA: 54283 MW: 09063C013DAE39C3 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 495;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 303 EPIYA 307

RESULT 10
ID GTR3_HUMAN STANDARD; PRT: 496 AA.
AC P11169; O9UG15;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Solute carrier family 2, facilitated glucose transporter, member 3
DE (Glucose transporter type 3, brain).
GN SLC2A3 OR GLUT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=89008414; PubMed=3170580;
RA Kayano T., Fukumoto H., Eddy R.L., Fan Y.-S., Byers M.G., Shows T.B.,
RA Bell G.I.;
RT "Evidence for a family of human glucose transporter-like proteins.
RT Sequence and gene localization of a protein expressed in fetal
RT skeletal muscle and other tissues."
RL J. Biol. Chem. 263:15245-15248(1988).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA Stuart C.A., Wen K.G., Acosta M., Wood T.G.;
RT "Resistance and expression of glucose transporters in human skeletal
RT muscle.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 288-496 FROM N.A.
RT TISSUE=Fetal brain;
RA Blum H., Bauersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 302-453 FROM N.A.
RT TISSUE=Articular cartilage;
RA Neama G., Richardson S., Bell S., Carter S., Mobasher A.;
RT "Molecular characterization and cloning of glucose transporters in
RT human articular chondrocytes.";
RN [5]
RP Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=98138446; PubMed=9477959;
RA Seatter M.J., de la Rue S.A., Porter L.M., Gould G.W.;
RT "QLS motif in transmembrane helix VII of the glucose transporter
RT family interacts with the C-1 (position) of D-glucose and is involved
RT in substrate selection at the exofacial binding site.";
RL Biochemistry 37:1322-1326(1998).
CC -!- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. PROBABLY A NEURONAL
CC GLUCOSE TRANSPORTER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC TRANSPORTER SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M20681; AABG1083.1; -
DR EMBL; AF274892; AAF82116.1; -
DR EMBL; AF274889; AAF82116.1; JOINED.
DR EMBL; AF274890; AAF82116.1; JOINED.
DR EMBL; AF274891; AAF82116.1; JOINED.
DR EMBL; ALI10298; CAB53739.1; -
DR EMBL; AY034634; AAK56796.1; -
DR PIR; A31986; A31986.
DR Genew; HGNC:11007; SLC2A3.
DR MIM; 138170; -
DR InterPro: IPR003663; CHO_transport.
DR InterPro: IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 2.
DR PRINTS; PR00171; SUGRTNSPORT.
DR TIGRFAMS; TIGR00879; SP; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane; Sugar transport; Transport; Glycoprotein;
KW Multigene family.
FT DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 6 26 1 (POTENTIAL).
FT DOMAIN 27 64 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 65 85 2 (POTENTIAL).
FT DOMAIN 86 93 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 94 114 3 (POTENTIAL).
FT DOMAIN 115 124 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 125 145 4 (POTENTIAL).
FT DOMAIN 146 153 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 154 174 5 (POTENTIAL).
FT DOMAIN 175 183 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 184 204 6 (POTENTIAL).
FT DOMAIN 205 269 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 270 290 7 (POTENTIAL).
FT DOMAIN 291 304 EXTRACELLULAR (POTENTIAL).

```

FT TRANSMEM 305 325 8 (POTENTIAL).
 FT DOMAIN 326 334 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 335 355 9 (POTENTIAL).
 FT DOMAIN 356 368 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 369 389 10 (POTENTIAL).
 FT DOMAIN 390 399 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 400 420 11 (POTENTIAL).
 FT DOMAIN 421 427 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 428 448 12 (POTENTIAL).
 FT DOMAIN 449 496 DEFINES SUBSTRATE SPECIFICITY (BY
 FT SITE 277 SIMILARITY).
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 303 303 E -> Q (IN REF. 3).
 FT CONFLICT 310 310 G -> S (IN REF. 3).
 FT CONFLICT 320 321 DV -> LL (IN REF. 3).
 FT CONFLICT 357 358 VN -> NH (IN REF. 3).
 FT CONFLICT 376 376 F -> C (IN REF. 3).
 FT CONFLICT 425 425 H -> Y (IN REF. 3).
 FT CONFLICT 432 432 H -> L (IN REF. 3).
 FT CONFLICT 482 482 E -> G (IN REF. 3).
 FT CONFLICT 482 482 E -> G (IN REF. 3).
 SQ SEQUENCE 496 AA: 53924 MW; F601CD6892F16516 CRC64;
 Query Match 100.08; Score 27; DB 1; Length 496;
 Best Local Similarity 100.08; Pred. No. 43;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EPIYA 5
 Db 303 EPIYA 307
 RESULT 11
 ID CATA_DROME STANDARD; PRT; 506 AA.
 AC P17336; OSVvml;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Catalase (EC 1.11.1.6).
 GN CAT OR CG6871.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96239139; PubMed=8660653;
 RA Orr W.C., Orr E.C., Legan S.K., Sohal R.S.;
 RT "Molecular analysis of the Drosophila catalase gene.";
 RL Arch. Biochem. Biophys. 330:251-258(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90301508; PubMed=2362827;
 RA Orr E.C., Bewley G.C., Orr W.C.;
 RT "cDNA and deduced amino acid sequence of Drosophila catalase.";
 RL Nucleic Acids Res. 18:3663-3663(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mixlos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.N., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE OF 76-92;
 RC STRAIN=Vallecas; TISSUE=wing imaginal disk;
 RX MEDLINE=93272852; PubMed=8500545;
 RA Santaren J.F., van Damme J., Puype M., Vandekerckhove J.,
 RA Garcia-Bellido A.;
 RT "Identification of Drosophila wing imaginal disc proteins by two-
 dimensional gel analysis and microsequencing.";
 RL Exp. Cell Res. 206:220-226(1993).
 CC -!- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
 CC SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
 CC PEROXIDE.
 CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
 CC -!- COFACTOR: HEME GROUP.
 CC -!- SUBUNIT: HOMOTETRAMER.
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.
 CC -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U00145; AAC13738.1;
 DR EMBL; X52286; CAA36529.1; ALT_SEQ.
 DR EMBL; AE003519; AAF49228.1; -.
 DR PIR; S12725; CSFF.
 DR HSSP; P04040; 1F4J.
 DR FlyBase; FBgn0000261; Cat.
 DR InterPro; IPRO02226; Catalase.
 DR Pfam; PF00199; catalase; 1.
 DR PRINTS; PR00067; CATALASE.
 DR PRODom; PR000510; Catalase; 1.
 DR PROSITE; PS00342; MICROBODIES_CTER; 1.
 DR PROSITE; PS00437; CATALASE_1; 1.
 DR PROSITE; PS00438; CATALASE_2; 1.
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
 KW Peroxisome.
 FT ACT_SITE 73 73 BY SIMILARITY.
 FT ACT_SITE 146 146 BY SIMILARITY.
 FT BINDING 356 356 PROXIMAL HEME LIGAND (BY SIMILARITY).
 FT SITE 504 506 MICROBODY TARGETING SIGNAL (POTENTIAL).
 FT CONFLICT 5 5 D -> Y (IN REF. 1).

```

FT  CONFLICT 12 12 I -> N (IN REF. 1).
SQ  SEQUENCE 506 AA; 57149 MW; 396377DC5F784ECE CRC64;

Query Match
Best Local Similarity 100.0%; Score 27; DB 1; Length 506;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 226 EPIYA 230

RESULT 12
ID HEX_ADE09 STANDARD; PRT; 512 AA.
AC P36853;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hexon protein (late protein 2) (fragment).
GN P11.
OS Human adenovirus type 9.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10527;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate Hicks;
RA Pring-Akerblom P.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Eiz B.;
RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE STRUCTURAL PROTEINS IN THE
CC VIRAL COAT AND IS SYNTHESIZED DURING LATE INFECTION.
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74664; CAA52728.1; -
DR EMBL; X74657; CAA52721.1; -
DR PIR; S37221; S37221.
DR PIR; S37279; S37279.
DR HSP; P03277; IDHX.
DR InterPro; IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
KW Coat protein; Hexon protein; Late protein.
FT NON_TER 1 1
FT NON_TER 512 512
SQ SEQUENCE 512 AA; 57119 MW; 87B7AF2A7452CF10 CRC64;

Query Match
Best Local Similarity 100.0%; Score 27; DB 1; Length 512;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 79 EPIYA 83

RESULT 13
ID APEL_SULTO STANDARD; PRT; 786 AA.
AC Q96279;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable aminopeptidase 1 (EC 3.4.11.-).
GN APE1 OR ST1746.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10345 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudo H. Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophillic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
CC -1- COFACTOR: Binds one zinc ion (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000987; BAB66834.1; -
DR InterPro; IPR001930; Ala_peptase.
DR InterPro; IPR000130; Zn_MTPeptase.
DR Pfam; PF01433; Peptidase_M1.1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metallopeptase; Aminopeptidase; Zinc; Complete proteome.
FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 271 271 BY SIMILARITY.
FT METAL 274 274 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 293 293 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 356 356 PROTON DONOR (POTENTIAL).
SQ SEQUENCE 786 AA; 90739 MW; 0CE5B6CD13CEC2D3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 27; DB 1; Length 786;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 103 EPIYA 107

RESULT 14
ID CGAL_HELPY STANDARD; PRT; 1147 AA.
AC P80200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN CAGA OR CAI.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petracca R., Burrone D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N.,

```



```

RA Rappuoli R.;
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal
RT ulcer.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=97121442; PubMed=8962108;
RA Censini S., Lange C., Xiang Z., Crabtree J., Ghiara P.,
RA Borodovsky M., Rappuoli R., Covacci A.;
RT "Cag, a pathogenicity island of Helicobacter pylori, encodes type I-
RT specific and disease-associated virulence factors.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).
RN [3]
RP SEQUENCE OF 181-190; 314-328; 366-377; 420-430; 494-506; 661-677;
RP 900-914 AND 1062-1077.
RA Herrmann V., Herrmann J., Kist M.;
RA Submitted (APR-1993) to the SWISS-PROT data bank.
CC -1- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70039; CAA49633.1; -
DR EMBL; AF282853; AAC44706.1; -
DR InterPro; IPR005169; CagA.
DR Pfam; PF03507; IVSec_cagA.
DR PRINTS; PR01553; TYPE4SSCAGA.
DR DOMAIN 880 885 POLY-ASN.
FT CONFLICT 320 320 G -> A (IN REF. 3).
FT CONFLICT 325 325 P -> F (IN REF. 3).
FT CONFLICT 328 328 R -> K (IN REF. 3).
FT CONFLICT 426 426 K -> E (IN REF. 3).
FT CONFLICT 429 429 Q -> E (IN REF. 3).
FT CONFLICT 673 675 AQN -> TED (IN REF. 3).
FT CONFLICT 901 901 A -> T (IN REF. 3).
FT CONFLICT 903 903 Q -> E (IN REF. 3).
FT CONFLICT 907 907 L -> P (IN REF. 3).
FT CONFLICT 910 910 P -> R (IN REF. 3).
FT CONFLICT 914 914 Q -> E (IN REF. 3).
FT CONFLICT 1072 1072 P -> S (IN REF. 3).
FT CONFLICT 1074 1074 S -> D (IN REF. 3).
SQ SEQUENCE 1147 AA; 128013 MW; AB92770835F68490 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 1147;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 890 EPIYA 894

RESULT 15
CAGA_HELPJ STANDARD; PRT; 1167 AA.
AC CAGA_HELPJ
AT Q9ZLTI;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE cytotoxicity associated immunodominant antigen (120 kDa protein)
DE (CAG pathogenicity island protein 26).
DE CAGA OR CAI OR CAG26 OR JHP0495.
OS Helicobacter pylori J99 (Campylobacter pylori J99).

```

```

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001483; AAD06073.1; -
DR InterPro; IPR005169; CagA.
DR InterPro; IPR004355; IVSec_cagA.
DR Pfam; PF03507; CagA; 1.
DR PRINTS; PR01553; TYPE4SSCAGA.
KW Antigen; Complete proteome.
FT DOMAIN 246 249 POLY-THR.
FT DOMAIN 882 889 POLY-ASN.
SQ SEQUENCE 1167 AA; 129729 MW; FD5E86B81CEBD0F2 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 950 EPIYA 954

Search completed: May 12, 2003, 11:53:50
Job time : 2.13248 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:47:39 ; Search time 0.42735 Seconds
(without alignments)
2410.751 Million cell updates/sec

Title: US-09-360-685B-10
Perfect score: 27
Sequence: 1 EPIYA 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvrius.*
- 16: sp_bacteriap.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	27	100.0	68	2	Q48256	Q48256 helicobacte
2	27	100.0	80	17	Q9HQ71	Q9hq71 halobacteri
3	27	100.0	115	11	Q9CQ44	Q9cq44 mus musculu
4	27	100.0	171	2	Q9F6N5	Q9f6n5 helicobacte
5	27	100.0	171	2	Q9F6M5	Q9f6m5 helicobacte
6	27	100.0	171	2	Q8RRN3	Q8rrn3 helicobacte
7	27	100.0	174	2	Q8RRN7	Q8rrn7 helicobacte
8	27	100.0	176	2	Q93A83	Q93a83 helicobacte
9	27	100.0	180	2	Q9F6Q2	Q9f6q2 helicobacte
10	27	100.0	188	2	Q9F6P5	Q9f6p5 helicobacte
11	27	100.0	189	2	Q9F6P7	Q9f6p7 helicobacte
12	27	100.0	189	2	Q9F6N9	Q9f6n9 helicobacte
13	27	100.0	189	2	Q9F6N7	Q9f6n7 helicobacte
14	27	100.0	189	2	Q9F6N2	Q9f6n2 helicobacte
15	27	100.0	189	2	Q9F6N1	Q9f6n1 helicobacte
16	27	100.0	189	2	Q9F6M9	Q9f6m9 helicobacte

17	27	100.0	190	2	Q9F6Q3	Q9f6q3 helicobacte
18	27	100.0	190	2	Q9F6Q0	Q9f6q0 helicobacte
19	27	100.0	190	2	Q9F6P6	Q9f6p6 helicobacte
20	27	100.0	190	2	Q9F6P1	Q9f6p1 helicobacte
21	27	100.0	191	2	Q9F6M7	Q9f6m7 helicobacte
22	27	100.0	192	2	Q9F6Q6	Q9f6q6 helicobacte
23	27	100.0	192	2	Q9F6P9	Q9f6p9 helicobacte
24	27	100.0	192	2	Q9F6P4	Q9f6p4 helicobacte
25	27	100.0	192	2	Q9F6P2	Q9f6p2 helicobacte
26	27	100.0	192	2	Q9F6M8	Q9f6m8 helicobacte
27	27	100.0	192	2	Q9F6M6	Q9f6m6 helicobacte
28	27	100.0	196	2	Q8RRN2	Q8rrn2 helicobacte
29	27	100.0	198	2	Q8RRU1	Q8rru1 helicobacte
30	27	100.0	198	2	Q8RRT5	Q8rrt5 helicobacte
31	27	100.0	198	2	Q8RRR0	Q8rrr0 helicobacte
32	27	100.0	199	2	Q8RRU2	Q8rru2 helicobacte
33	27	100.0	200	2	Q8RRU4	Q8rru4 helicobacte
34	27	100.0	200	2	Q8RRU0	Q8rru0 helicobacte
35	27	100.0	200	2	Q8RRT0	Q8rrt0 helicobacte
36	27	100.0	201	2	Q8RRU5	Q8rru5 helicobacte
37	27	100.0	201	2	Q8RRU3	Q8rru3 helicobacte
38	27	100.0	201	2	Q8RRT7	Q8rrt7 helicobacte
39	27	100.0	203	2	Q8RRQ8	Q8rrq8 helicobacte
40	27	100.0	203	2	Q8RRQ7	Q8rrq7 helicobacte
41	27	100.0	203	2	Q8RRQ6	Q8rrq6 helicobacte
42	27	100.0	204	2	Q8RRT3	Q8rrt3 helicobacte
43	27	100.0	204	2	Q8RRS5	Q8rrs5 helicobacte
44	27	100.0	205	2	Q8RRT9	Q8rrt9 helicobacte
45	27	100.0	205	2	Q8RRT8	Q8rrt8 helicobacte

ALIGNMENTS

RESULT 1

Q48256 ID Q48256 PRELIMINARY; PRT; 68 AA.
AC Q48256;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cai protein (Fragment).
GN CAI.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G39;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petracca R., Burroni D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N., Rappuoli R.;
RT "Molecular characterization of the 128 kDa immunodominant antigen of
RL helicobacter pylori associated with cytotoxicity and duodenal ulcer.";
Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
DR EMBL; X70038; CAA49632.1; -;
DR InterPro; IPR005169; CagA.
DR Pfam; PF03507; CagA; 2.
FT NON_TER 1
FT NON_TER 68
SQ SEQUENCE 68 AA; 7428 MW; D66734B89840C7F2 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
|
|
|
|
Db 41 EPIYA 45

RESULT 2

```

Q9HQ71      PRELIMINARY;      PRT;      80 AA.
ID Q9HQ71;
AC Q9HQ71;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE Vng1300H.
GN Vng1300H.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL PROC. Natl Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005053; AAG19646.1; -.
SQ SEQUENCE 80 AA; 8810 MW; 410A92DF96257423 CRC64;

Query Match 100.0%; Score 27; DB 17; Length 80;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 47 EPIYA 51

RESULT 3
ID Q9CQ44 PRELIMINARY; PRT; 115 AA.
AC Q9CQ44;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE 181002LJ13RIK protein.
GN 181002LJ13RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=STOMACH, AND BRAIN;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staib F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008902; BAB25961.1; -.
DR EMBL; AK003010; BAB22510.1; -.
DR MGD; MGI:1913529; 181002LJ13RIK.
SQ SEQUENCE 115 AA; 12464 MW; C9165E64EBDE3AB9 CRC64;

Query Match 100.0%; Score 27; DB 11; Length 115;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 105 EPIYA 109

RESULT 4
ID Q9F6N5 PRELIMINARY; PRT; 171 AA.
AC Q9F6N5;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (Fragment).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR50;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289454; AAG09903.1; -.
DR InterPro; IPR005169; CagaA.
DR Pfam; PF03507; CagaA; 1.
FT NON_TER 1
FT NON_TER 171
SQ SEQUENCE 171 AA; 18808 MW; 86670327F2053BA7 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 82 EPIYA 86

RESULT 5
ID Q9F6M5 PRELIMINARY; PRT; 171 AA.
AC Q9F6M5;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (Fragment).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR61;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,

```

```

RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289464; AAG09913.1;
DR InterPro; IPR005169; CagA.
DR Pfam; PF03507; CagA; 1.
FT NON_TER 171 171
SQ SEQUENCE 171 AA; 18850 MW; A21EE947EDI5283D CRC64;

Query Match 100.0%; Score 27; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 82 EPIYA 86

RESULT 6
Q8RRN3 ID Q8RRN3 PRELIMINARY; PRT; 171 AA.
AC Q8RRN3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein CagA (Fragment).
GN CAGA.
OS Helicobacter pylori.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLOMBIA82;
RA Yamaoka Y., Gutierrez O., Saitou N., Kodama T., Kim J.G., Kashima K.,
RA Ramirez F.C., Mahachai V., Osato M.S., Graham D.Y.;
RT "Helicobacter pylori in North and South America before Columbus:
RT Studies of human migration.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057104; BAB87471.1;
DR InterPro; IPR005169; CagA; 1.
FT NON_TER 171 171
SQ SEQUENCE 171 AA; 18247 MW; 2FDBD3D8804E4735 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 124 EPIYA 128

RESULT 7
Q8RRN7 ID Q8RRN7 PRELIMINARY; PRT; 174 AA.
AC Q8RRN7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein CagA (Fragment).
GN CAGA.
OS Helicobacter pylori.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-USN2754;
RA Yamaoka Y., Gutierrez O., Saitou N., Kodama T., Kim J.G., Kashima K.,

```

```

RA Ramirez F.C., Mahachai V., Osato M.S., Graham D.Y.;
RT "Helicobacter pylori in North and South America before Columbus:
RT Studies of human migration.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057100; BAB87467.1;
DR InterPro; IPR005169; CagA; 1.
FT NON_TER 174 174
SQ SEQUENCE 174 AA; 18851 MW; E3944CC895A5E558 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 86 EPIYA 90

RESULT 8
Q93A83 ID Q93A83 PRELIMINARY; PRT; 176 AA.
AC Q93A83;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin-associated gene A (Fragment).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1091;
RA Dong O., O'Morain C.A., Buckley M.;
RT "A new variable segment in the CagA 3' region of Helicobacter
RT pylori.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF427098; AAU28091.1;
DR InterPro; IPR005169; CagA.
DR Pfam; PF03507; CagA; 1.
FT NON_TER 176 176
SQ SEQUENCE 176 AA; 19316 MW; C218EDA09179A681 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EPIYA 5
Db 82 EPIYA 86

RESULT 9
Q9F6Q2 ID Q9F6Q2 PRELIMINARY; PRT; 180 AA.
AC Q9F6Q2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (Fragment).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR7;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;

```

```

RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL: AF289437; AAG09886.1;
DR InterPro: IPR005169; Caga.
DR Pfam: PF03507; Caga; 1.
FT NON_TER 180 180
FT NON_TER 180 180
SQ SEQUENCE 180 AA; 992ED95B90FFFC44 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 100 EPIYA 104

RESULT 10
Q9F6P5 PRELIMINARY; PRT; 188 AA.
AC Q9F6P5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (fragment).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR21;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL: AF289444; AAG09893.1;
DR InterPro: IPR005169; Caga.
DR Pfam: PF03507; Caga; 1.
FT NON_TER 188 188
FT NON_TER 188 188
SQ SEQUENCE 188 AA; 128E2C60BDFC67BF CRC64;

Query Match 100.0%; Score 27; DB 2; Length 188;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 99 EPIYA 103

RESULT 11
Q9F6P7 PRELIMINARY; PRT; 189 AA.
AC Q9F6P7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (fragment)..
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=CR19;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL: AF289442; AAG09891.1;
DR InterPro: IPR005169; Caga.
DR Pfam: PF03507; Caga; 1.
FT NON_TER 189 189
FT NON_TER 189 189
SQ SEQUENCE 189 AA; 20758 MW; 39E55929869BCF3 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 100 EPIYA 104

RESULT 12
Q9F6N9 PRELIMINARY; PRT; 189 AA.
AC Q9F6N9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (fragment).
GN CAGA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR42;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL: AF289450; AAG09899.1;
DR InterPro: IPR005169; Caga.
DR Pfam: PF03507; Caga; 1.
FT NON_TER 189 189
FT NON_TER 189 189
SQ SEQUENCE 189 AA; 20508 MW; 15C59E08B63BB5EE CRC64;

Query Match 100.0%; Score 27; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
Db 100 EPIYA 104

RESULT 13
Q9F6N7 PRELIMINARY; PRT; 189 AA.
AC Q9F6N7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (fragment).
GN CAGA.

```

```

OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR48;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289452; AAG09901.1; -
DR InterPro; IPR005169; CagA.
DR Pfam; PF03507; CagA; 1.
FT NON_TER 1
FT NON_TER 189
FT NON_TER 189
SQ SEQUENCE 189 AA; 20521 MW; 4CD371C3D1201D1F CRC64;

Query Match 100.0%; Score 27; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 100 EPIYA 104

RESULT 14
Q9F6N2 PRELIMINARY; PRT; 189 AA.
AC Q9F6N2;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (Fragment).
GN CAG.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR53;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289457; AAG09906.1; -
DR InterPro; IPR005169; CagA.
DR Pfam; PF03507; CagA; 1.
FT NON_TER 1
FT NON_TER 189
FT NON_TER 189
SQ SEQUENCE 189 AA; 20541 MW; AFD7F997E9148508 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 100 EPIYA 104

RESULT 15
Q9F6N1 PRELIMINARY; PRT; 189 AA.
AC Q9F6N1;

```

```

DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Cytotoxin associated protein A (Fragment).
GN CAG.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR54;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289458; AAG09907.1; -
DR InterPro; IPR005169; CagA.
DR Pfam; PF03507; CagA; 1.
FT NON_TER 1
FT NON_TER 189
FT NON_TER 189
SQ SEQUENCE 189 AA; 20602 MW; A023D7E7A7919588 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EPIYA 5
DB 103 EPIYA 107

Search completed: May 12, 2003, 11:55:24
Job time : 1.67735 secs

```

THIS PAGE BLANK (USPTO)